

Nama : Adib Annur Rohman

NIM : A11.2021.13712

Kelas : A11.4404

Tugas Akhir Data Mining
Klasifikasi Penyakit Diabetes di Kalangan Masyarakat
Menggunakan Algoritma Decision Tree

Dataset yang digunakan merupakan dataset Diabetes yang berasal dari Kaggle. Dataset ini original dari the N. Inst. Of Diabetes & Diges. & Kidney Dis. Dataset ini berisi sampel-sampel dengan total 769 sampel data, dan ada juga beberapa atribut seperti dibawah ini :

- Kehamilan : Berapa kali hamil
- Glukosa : Konsentrasi glukosa plasma selama 2 jam dalam tes toleransi glukosa oral
- Tekanan Darah : Tekanan darah diastolic (mm Hg)
- Ketebalan Kulit : Ketebalan lipatan kulit trisep (mm)
- Insulin : Insulin serum 2 jam (mu U/ml)
- BMI : Indeks massa tubuh (berat dalam kg/(tinggi dalam m)²)
- Silsilah Diabetes : Fungsi silsilah diabetes
- Usia : Usia (tahun)
- Hasil : Variabel kelas (0 atau 1)

Tahapan Eksperimen :

- Pertama kita persiapkan dataset nya yang sudah kita dapatkan dengan jumlah 769 data sampel dan 9 atribut.
- Variabel outcome dikeluarkan disini yang berisi 768 rows dan 8 column.
- Kemudian saya lakukan pemodelan dengan menggunakan Decision Tree.
- Pola yang dihasilkan bisa berbentuk tree atau if-then.
- Setelah itu saya evaluasi menggunakan skor yang didapatkan.
- Mengevaluasi menggunakan matriks confusion.
- Mengevaluasi menggunakan laporan klasifikasi.
- Memeriksa nilai prediksi.
- Mengimport modul untuk memvisualisasikan pohon keputusan.
- Hasil pola dari data berupa decision tree (pohon keputusan).
- Setelah keluar hasil pohon keputusan kemudian saya evaluasi lagi dengan visualisasi pohon keputusan diatas agar lebih baik.
- Dan atribut atau factor semua berpengaruh terhadap hasil akurasi.
- Kemudian saya membuat program peningkatan Kesehatan untuk mengurangi masalah penyakit diabetes dikalangan masyarakat.